Journal of Clinical and Translational Science

www.cambridge.org/cts

Operationalizing community engagement evaluation: A structured and scalable approach using the RE-AIM framework and net effects diagrams

Brian Do-Golden[1], Nicole Wolfe[1], Nicole M.G. Maccalla[1], James Settles[1] and Michele D. Kipke[1,2,3]

[1]University of Southern California, Southern California Clinical and Translational Science Institute, Los Angeles, CA, US; [2]University of Southern California, Office of the Senior Vice President of Health Affairs, Los Angeles, CA, US and [3]Children's Hospital of Los Angeles, Department of Pediatrics Los Angeles, CA, US

Research Article

Cite this article: Do-Golden B, Wolfe N, Maccalla NMG, Settles J, and Kipke MD. Operationalizing community engagement evaluation: A structured and scalable approach using the RE-AIM framework and net effects diagrams. *Journal of Clinical and Translational Science* **9**: e255, 1–12. doi: 10.1017/cts.2025.10190

Received: 28 March 2025
Revised: 24 September 2025
Accepted: 23 October 2025

Keywords:
Community engagement (CE); implementation science (IS); re-aim framework; evidence-based program adaptation; community development evaluation

Corresponding author:
N. Wolfe; Email: nicole.wolfe@med.usc.edu

Abstract

Introduction: Community engagement (CE) is essential in Clinical and Translational Science (CTS), yet its evaluation remains inconsistent and often lacks standardization. The RE-AIM framework (Reach, Effectiveness, Adoption, Implementation, Maintenance) offers a promising structure for evaluating CE efforts, but its application in dynamic, community-based contexts is often limited by data variability and implementation complexity. *Methods:* We developed and applied a seven-step, structured, and replicable approach to operationalizing RE-AIM for program evaluation. This method includes the use of tailored RE-AIM subdomains, standardized scoring systems, and visual analytics through Net Effects Diagrams. *Results:* We applied this framework to our community-based health education workshops delivered in English and Spanish across Los Angeles, using participant surveys and facilitator feedback data. The operationalized framework enabled consistent assessment and comparison between language groups. Spanish-language workshops outperformed English-language workshops (ELWs) in measures of attendance, participant satisfaction, and short-term effectiveness. Visualizations using Net Effects Diagrams facilitated collaboration among stakeholders to interpret program outputs and outcomes, supporting actionable insights for program adaptation. Differences between workshop groups will inform changes to recruitment and content delivery strategies in ELWs. *Conclusions:* This approach offers a transparent, scalable, and context-sensitive method for assessing CE programs. It supports data-driven decision-making, continuous program improvement, and stakeholder engagement. While developed for CE initiatives, the method is broadly adaptable to other community and public health programs. Future efforts will include expanded outcome tracking, integration into dashboards, and dissemination as a toolkit for broader adoption within and beyond the CTS Award network.

© The Author(s), 2025. Published by Cambridge University Press on behalf of Association for Clinical and Translational Science. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Introduction

The Clinical and Translational Science Award (CTSA) Program in the United States, funded by the National Institutes of Health, supports the advancement of efficient research translation, speeding up the process of taking the discoveries made in laboratories, clinics, and communities to inform and improve processes and policies [1]. Community engagement (CE) programs are an essential and required element of each of the 60+ CTSA hubs, recognizing that meaningful engagement with communities enhances research relevance, accessibility, and impact [2]. CE is central to CTS, particularly for building and maintaining trust in populations that have historically lower rates of participating in research and accessing the healthcare system [1].

The CE Core within the Southern California Clinical and Translational Science Institute (SC CTSI) serves as a critical bridge between the community, researchers, and academic institutions. We focus on fostering meaningful engagement with the communities in South, Central, and the Eastside of Los Angeles, communities that have long faced limited access to healthcare, disproportionately higher rates of chronic disease, and lower rates of participation in clinical research.

The SC CTSI CE Core implements a range of initiatives designed to promote health education, research literacy, and bidirectional collaboration between community members and academic researchers. One key component is our education and training program, which delivers in-person and virtual workshops on health and research-related topics that aim to impact community well-being, including nutrition, mental health, diabetes, and other chronic diseases. Additionally, we offer a well-rounded research training program tailored for

community health workers/promotores de salud, equipping them with the knowledge and skills to support health research and outreach efforts in their communities. Beyond community education, we also provide consultations to researchers, guiding them in integrating CE frameworks and community-based participatory research methodologies into their work.

CE efforts are vital for addressing differences in health outcomes and developing and maintaining key partnerships that bridge the gap between scientific discovery and the well-being of communities [3]. Despite its importance, evaluating CE efforts remains challenging due to the lack of robust, standardized, replicable frameworks. This makes it difficult to identify best practices, compare outcomes, and maintain effective engagement approaches [4]. Traditionally, the evaluation of CE work often relies on academic outcomes such as publications rather than assessing the real-world impact of CE activities on communities, partnerships, and translational research outcomes [5]. Therefore, a robust, standardized evaluation framework is needed that encapsulates the *Reach* and *Effectiveness* of CE initiatives as well as their *Adoption*, *Implementation*, and *Maintenance*.

Given the complexity and variability of CE activities, an evaluation framework must assess impact at multiple levels and offer insights that facilitate practical programmatic decision-making. Implementation science (IS) has been used for CE evaluation to provide a coordinated structure for assessing real-world implementation and identifying components that lead to success [6]. Traditionally focusing on shrinking the gap between research discoveries and their practical application in community and clinical settings, IS has also been used to improve adoption and effectiveness of evidence-based practices in behavior interventions, healthcare delivery, and public health [7,8]. Within the CTS context, IS principles have accelerated the dissemination and implementation of research findings, translating clinical advancements into meaningful community and population-level health improvements [9]. Despite its integration into CTSA hubs' planning and implementation, IS frameworks remain underutilized for evaluating CTSA hubs' CE initiatives, limiting the ability to fully assess their impact [10].

Among IS frameworks, RE-AIM allows for the adaptability of community-based research and its ability to capture implementation, engagement, and effectiveness outcomes in a standardized manner. This has proven to be valuable in evaluating CE initiatives by systematically describing key dimensions of engagement and impact [11,12]. Originally developed for evaluating public health and clinical interventions, RE-AIM has been increasingly recognized as a useful tool for community-based research, as it allows for the examination of both individual- and organizational-level factors influencing program success [12]. Partnering visualization tools with the RE-AIM framework could aid in the communication of a multitude of complex CE program outputs and outcomes, which can make it a promising application for evaluating CE activities with CTSAs [13]. Therefore, RE-AIM provides the structure and level of flexibility that assesses individual, participant level, and organizational programmatic factors, making it a good fit for CE initiatives [12].

Evaluating CE programs often involves working with multiple data types collected across varying contexts, making it challenging to quickly monitor and interpret programmatic effects. Visual tools can play a role in addressing these difficulties by helping evaluators compare outputs/outcomes across domains, populations, and settings. Radar charts, what we refer to as "Net Effects Diagrams" in this paper, offer a way to display multivariate data in a way that is both accessible and actionable [14]. When paired with standardized scoring methods, these visualizations can provide both summative insights (i.e., 'Does it work?,' "Is it effective?") and formative feedback that supports program planning, monitoring, and improvement.

This paper outlines a generalizable method for operationalizing programmatic evaluation to produce and monitor outcomes, using our community-based educational workshops as an illustrative case in the results section. This structured, step-by-step approach, centered around the RE-AIM framework and supported by a standardized scoring system, is designed to be transparent and replicable across diverse program contexts and settings. This method is a structural proof-of-concept that can be readily adapted for unique contexts (e.g., setting, program, organizational structure, program purpose, etc.) outside of the example provided in this paper.

Methods

This methods section outlines a seven-step, structured process for operationalizing RE-AIM, through standardized scoring and Net Effects Diagrams, to support customizable, scalable, and replicable programmatic and context specific evaluation (Figure 1).

Step 1: Identify program and operationalize RE-AIM domains

The first step involves selecting the program to be evaluated and defining each RE-AIM domain in operational terms. Programs should be selected based on their strategic relevance, implementation maturity, and availability of evaluation data. Each domain should be interpreted in a way that aligns with the program's goals, intended impact, and contextual factors (e.g., setting, population served, or implementation approach). For example, *Reach* may refer to the number of individuals engaged, proportion of the target population engaged, and/or degree of demographic alignment between participants and intended audiences. These definitions should be developed collaboratively with program staff and key stakeholders to ensure alignment with both implementation goals and community priorities.

Step 2: Determine RE-AIM subdomains for evaluation

Domains can be further divided into relevant subdomains to increase specificity and enhance the actionability of findings. Subdomains allow evaluators to focus on meaningful components of each domain that better reflect implementation realities and measurement priorities. Some examples include:

- Reach: frequency of programming, participant attendance, representativeness of attendees
- Effectiveness: increases in participant knowledge, satisfaction, and self-efficacy
- Implementation: fidelity to evidence-based practices, perceived responsiveness to community needs

The number of subdomains may vary depending on the domain and the program. Some may only require one subdomain, while others may benefit from multiple. Subdomains are not always necessary in every domain. For instance, *Maintenance* may be evaluated through sustained outcomes over time or program continuity across years and may not require further breakdown. Whereas the domain *Adoption* can be conceptualized as differences in the *Reach*, *Effectiveness*, and *Implementation* outputs

https://doi.org/10.1017/cts.2025.10190 Published online by Cambridge University Press

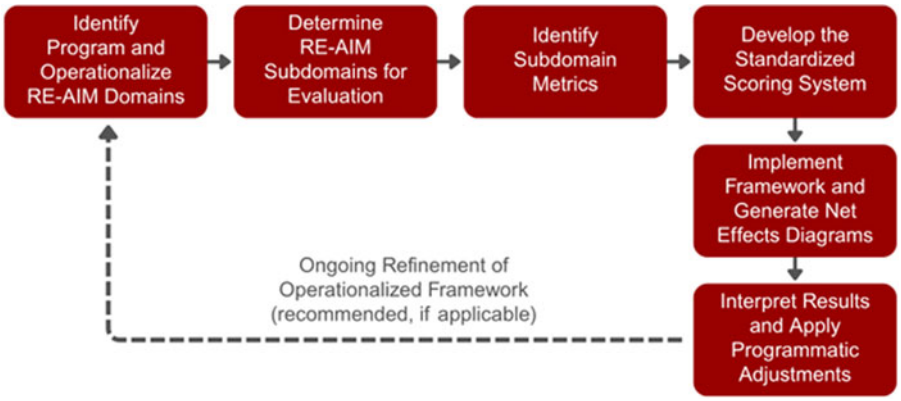


Figure 1. Seven-step process for operationalizing RE-AIM. RE-AIM, reach, effectiveness, adoption, implementation, and maintenance.

between contexts, such as site type, implementation strategy, or time frames.

Step 3: Identify subdomain metrics

For each subdomain, evaluators must identify one or more metrics to capture performance. These metrics should be relevant to program goals, interpretable by implementers and stakeholders, and consistently collected across time and groups. Metrics can include a variety of quantitative indicators, such as Likert-scale responses (i.e., satisfaction, confidence), pre/post assessments (i.e., knowledge gain), observational or staff-reported ratings (i.e., outreach efforts, participant engagement), and attendance or frequency counts.

When a single metric does not fully represent a subdomain, multiple indicators can be combined to reflect its broader scope. Metric identification should be grounded in the program's logic model or objectives, and can draw from existing tools, adapted instruments, or newly developed items that align with the evaluation and program's needs. Metrics can comprise various perspectives (e.g., participant and facilitator feedback) as a form of triangulation to increase credibility and validity in the findings by cross-verifying information.

Step 4: Develop the standardized scoring system

The choice of central tendency should reflect data distribution as means provide more granular differentiation but are sensitive to outliers, whereas medians resist skewness but may obscure meaningful variation. Outliers should also be assessed, as they can either represent meaningful real-world differences or distort values/outputs, and evaluators should document whether they are included, excluded, or analyzed separately. Reporting measures of spread such as standard deviation or interquartile range alongside any average scores used is recommended to provide additional context and avoid masking variability.

To support comparability and visualization across subdomains and domains, all selected metrics should be converted to a standardized scale. A common 1-to-5 scoring scale is recommended, where 1 indicates the lowest performance and 5 the highest. Standardization enables aggregation and supports intuitive visual interpretation. The scoring process generally involves:

1. Reviewing historical data or relevant benchmarks for each metric.
2. Drafting preliminary score thresholds (e.g., using quintiles or meaningful performance benchmarks).
3. Facilitating a review with subject matter experts or experienced program staff to validate cutoffs.
4. Adjusting thresholds based on contextual insights, ensuring the scores reflect real-world differences in performance.
5. Documenting the finalized scoring system along with the rationale for each range.

Once the general process is established, it can be applied to individual metrics to translate raw data into standardized scores. This step is crucial for ensuring that results from diverse measures can be interpreted on the same scale. It is important to note that scoring thresholds are not universal and should be tailored to the context of the program and refined over time as goals evolve or more data becomes available, for example as when measuring percent change (0.0–9.9% change = score of 1; 10.0–19.9% = 2; 20.0–29.9% = 3, etc). When applying this approach, evaluators should carefully consider how to summarize and compare data between groups.

Step 5: Implement framework and gnerate net effects diagrams

Standardized scores can now be calculated for each metric. For subdomains with multiple metrics, the individual standardized scores are averaged to generate a subdomain average score. Scores were visualized with Net Effects Diagrams to aid interpretation and enable comparisons across *Adoption* level variants (e.g., site type, implementation strategy, or time frame). These diagrams plot subdomain scores along axes extending from a central point. Each comparative program variant (e.g., between phases of a pilot program), group (between sites), or time frame (quarter vs. quarter) can be displayed on the chart, making differences easy to identify and interpret. These diagrams can be created using common tools like Microsoft Excel and are helpful in communicating patterns across programs thus informing programmatic assessment and decisions.

Step 6: Interpret results and apply programmatic adjustments

Teams should engage in regular review sessions to interpret Net Effects Diagrams and subdomain scores. Comparing results across RE-AIM domains can help uncover relationships (e.g., a high-reach program with low effectiveness) and inform adjustments. If

https://doi.org/10.1017/cts.2025.10190 Published online by Cambridge University Press

the program underperforms in any domain, examining subdomain results can clarify shortcomings and highlight variability that suggests inconsistent quality.

Discussions with program staff and stakeholders can help generate ideas of how to improve. Responsive adaptations could include modifying outreach strategies if *Reach* is low, updating content or delivery methods if satisfaction or knowledge scores are poor, or enhancing facilitator training if *Implementation* fidelity scores drop. Visualizing the outputs/outcomes supports transparency, shared understanding, and use of evaluation data in decision-making, which aids in communicating findings to a range of audiences.

Step 7: Ongoing refinement of operationalized framework

Evaluation frameworks should be viewed as dynamic tools that evolve alongside the programs they assess. As program goals shift, implementation strategies change, new staff is hired, or new audiences are engaged, the operationalized RE-AIM framework should be revisited to ensure continued alignment and relevance. Given the active participation of program staff in the development of this approach and the assessment of programmatic performance, efforts to "calibrate" staff observations and ensure a reasonable level of inter-rater reliability are required.

This step involves periodically reviewing and refining each element of the approach, including operational RE-AIM domain definitions, subdomains including how they are defined, selected metrics, and scoring thresholds. Programs that operate long term, scale to new settings, or undergo significant changes in content or implementation would particularly benefit from regular reassessment and ongoing calibration (e.g., annually or following major programmatic shifts). These reviews help ensure the framework remains responsive to changing conditions and continues to reflect both implementation goals and community values, while also allowing for benchmark comparisons over time. Any updates to the framework should be clearly documented and include the rationale for the change and its implications for longitudinal comparisons.

Stakeholder input, including that of program staff, participants, and community members, should inform these refinements. Qualitative insights or emergent themes from community feedback can highlight blind spots or unintended consequences in the evaluation approach that may not be captured through quantitative metrics alone. Incorporating ongoing refinement as a component of the approach enhances adaptability, ensures sustained alignment with program objectives, and supports long-term effectiveness of the evaluation system.

Results

This section presents the results of applying the seven-step approach to operationalize the RE-AIM evaluation framework by using our community-based health education workshops as an illustrative example. Workshops were delivered in both English and Spanish, providing an opportunity to assess program performance across language groups utilizing the standardized scoring system and generating Net Effects Diagrams. The survey data presented in this section were collected between January and September 2024 from a total of 81 workshops. There were 26 workshops delivered in English to a total of 215 participants, 134 of whom completed the pre and post surveys, and 55 workshops delivered in Spanish to a total of 1,103 participants, 725 of whom completed the pre and post surveys. It should be noted that not all who completed the surveys answered every question either intentionally or unintentionally, therefore there is variation in item-level response rates.

Step 1: Identify program and operationalize RE-AIM domain

The workshops were designed to improve participants' knowledge and confidence on health topics related to nutrition, chronic conditions, and mental health, among others. Workshops were tailored to and delivered across multiple sites. RE-AIM domains were defined for related CE activities in collaboration with program staff and leadership, and are as follows:

- Reach: The extent to which the intended population is engaged in programming and services*.
- Effectiveness: The extent to which programming and services* achieve the intended short-, medium-, and long-term goals as stated in the logic model.
- Adoption: The transferability and scalability of programming and services* to other settings (e.g., expand partnerships) and participants (e.g., providing activities in English and Spanish). The innovation and uptake of service models (e.g., place-based approaches) and innovations (e.g., dissemination efforts).
- Implementation: The extent to which evidence-based practices and community perspectives are used to shape programming and services* while testing innovative service and dissemination models.
- Maintenance: The extent to which programs, services, and service models are sustainable over time (e.g., partnerships, examples of institutionalization, additional funding sources, integration into existing institutions/programs/ leadership/ staffing, and/or changes in institutional policies/practices).

* *SC CTSI CE programming and services include, but are not limited to, educational offerings, connections to care, and clinical trial promotion and participation.*

Step 2: Determine RE-AIM subdomains for evaluation

Subdomains were then identified in collaboration with program staff and members from our leadership and executive teams to reflect the program's key implementation components, outcomes of interest, and available data sources (Table 1). These subdomains were meant to operationalize each RE-AIM domain in a way that would reflect workshop-specific design and implementation features.

Step 3: Identify subdomain metrics

Metrics were identified for each subdomain based on existing evaluation tools. Items were then added to existing data collections tools as needed to ensure each subdomain was supported by one or more relevant metrics. Pre-workshop surveys were administered to participants to collect demographic information and test their knowledge on the workshop topic (i.e., pre-workshop knowledge score). Post-workshop surveys were administered to participants to reassess their knowledge (i.e., post-workshop knowledge), along with questions to assess their self-efficacy/confidence levels, satisfaction, and intention to apply what they learned. Attendance sheets were used to log attendance. CE staff facilitating the workshops

https://doi.org/10.1017/cts.2025.10190 Published online by Cambridge University Press

Table 1. SC CTSI CE Operationalization of RE-AIM Subdomains

Domain: subdomain	Metric	Description
Reach: Frequency	Average monthly number of workshops offered	The average number of workshops offered per month.
Reach: Attendance	Perceived outreach effort by CE staff	CE staff perception rating of the relative effort made to do outreach for and promote the workshop.
	Average attendance per workshop	The average number of participants per workshop.
	Perceived participant turnout by CE staff	CE staff perception rating of the level of participant turnout to the workshop based on outreach and recruitment efforts.
	Characteristics of participants	CE staff perception rating of the degree to which the characteristics of the participant group of a workshop aligned with the expected characteristics based on the recruited for the workshop.
Effectiveness: Immediate Reaction	Participant satisfaction	The average participant rating on post-workshop participant satisfaction statements on a 4-point scale.
	Perceived level of participant engagement by CE staff	CE staff perception rating of the level of engagement by participants.
Effectiveness: Short-Term Goals	Participant change in knowledge	The average percent change in participant knowledge scores between pre-workshop and post-workshop.
	Participant perceived knowledge improvement	The average participant rating of their perceived knowledge improvement.
	Participant self-efficacy/ confidence levels	The average participant rating of their confidence levels in using what they learned and/or performing a skill learned.
	Participant intention to apply workshop content in their own lives	The average participant rating for their intention to incorporate what they learned from the workshop into their lives.
Effectiveness: Medium-Term Goals	Participant application of knowledge/skill	The average participant rating of how they have applied what they learned from a workshop.
	Participant impact	The average participant rating in how much the participation in a workshop has improved their health/lives.
Implementation: Evidence-Based	Level of evidence-based content rating	CE staff perception rating of the level of evidence-based information/ practices incorporated in the workshop.
Implementation: Responsiveness to Community Needs	Expert reflection/ staff perceptions - how did it go?	CE staff perception rating of how they felt about the workshop overall.
	Level of cultural relevancy	CE staff perception rating of how culturally relevant the workshop content and discussions were for the intended audience.

SC CTSI = Southern California Clinical and Translational Science Institute; CE = Community Engagement; RE-AIM = Reach, Effectiveness, Adoption, Implementation, and Maintenance.

completed a post-workshop facilitator feedback form where they rated statements related to their perceived understanding of outreach efforts, participant engagement, and effectiveness of the workshop. All data were collected and managed using REDCap electronic data capture tools hosted at Yale University [15,16]. An outline of the specific metrics for each subdomain can be found in Tables 2 and 3.

Step 4: Develop the standardized scoring system

We examined the distributions of data to determine appropriate measures of central tendency for comparing English- and Spanish-language groups. Given the small, variable sample sizes, we used mean values as the primary summary statistic, as they offered clearer group differences than medians, which in some cases (e.g., knowledge gain percent change) were uniformly 0.00% and thus uninformative. We did not exclude outliers from the central tendency analysis, as they represented meaningful participant experiences that we wanted to capture. Standard deviations are reported alongside mean values in Tables 2 and 3 to provide context about variability.

We then developed a standardized scoring system for each subdomain metric, allowing for consistent interpretation, comparison, and aggregation of results across workshops and between participant groups. Standardized scoring thresholds for each metric were developed iteratively, incorporating historical workshop data to identify performance distributions and input from senior staff. Using historical data, we generated preliminary thresholds based on percentile groupings or natural breakpoints in the data, while also being mindful of ceiling effect threats. For newly developed metrics, we used a combination of informed estimation and practical experience to determine initial scoring thresholds.

The historical data distributions, drafted thresholds, and potential case examples were then reviewed by senior staff. During this review process, they were asked to evaluate the thresholds and provide feedback based on the following criteria.

- Real-world interpretability: Do the thresholds meaningfully distinguish levels of program performance from an implementation perspective?
- Actionability: Would a score in a particular range prompt a different programmatic response or improvement strategy?

https://doi.org/10.1017/cts.2025.10190 Published online by Cambridge University Press

Table 2. Operationalized outputs for English-language workshops (January - September 2024)

RE-AIM Domain: Subdomain	Metric	Value (SD)	Standardized Scoring System	Standardized Metric Score	Subdomain Average Score
Reach: Frequency	Average monthly number of workshops offered	2.89 (3.73)	1 = 0.00-0.49 2 = 0.50-0.99 3 = 1.00-1.99 4 = 2.00-2.99 5 = 3.00+	4	4.0
Reach: Attendance	Perceived outreach effort (from CE staff)	4.54 (0.80)	1 = 1.00-2.99 2 = 3.00-3.49 3 = 3.50-3.99 4 = 4.00-4.49 5 = 4.50+	5	4.5
	Average attendance per workshop	8.27 (6.70)	1 = 0.00-1.99 2 = 2.00-4.99 3 = 5.00-7.99 4 = 8.00-11.99 5 = 12.00+	4	
	Perceived participant turnout (from CE staff)	3.81 (0.96)	1 = 1.00-2.59 2 = 2.60-3.19 3 = 3.20-3.79 4 = 3.80-4.39 5 = 4.40+	4	
	Characteristics of participants (actual in the context of recruited) (from CE staff)	4.50 (0.64)	1 = 1.00-2.99 2 = 3.00-3.49 3 = 3.50-3.99 4 = 4.00-4.49 5 = 4.50+	5	
Effectiveness: Immediate reaction	Participant satisfaction	3.56 (0.53)	1 = 1.00-2.49 2 = 2.50-3.09 3 = 3.10-3.39 4 = 3.40-3.69 5 = 3.70+	4	4.0
	Perceived level of participant engagement (from CE staff)	3.85 (0.91)	1 = 1.00-2.59 2 = 2.60-3.19 3 = 3.20-3.79 4 = 3.80-4.39 5 = 4.40+	4	

(*Continued*)

- Alignment with expectations: Do the thresholds reflect what staff would consider low, average, or high performance based on their field experience and program goals?

Based on this feedback, we adjusted thresholds to better reflect implementation realities and added narrative justifications for each score range to improve transparency and internal replicability. For example, for the participant change in knowledge metric, the threshold for a score of 3 was raised from 10 to 20% change after staff indicated that a 10% increase was commonly achieved and thus not representative of "average" metric performance. Similarly, satisfaction scores were reclassified so that a score of 5 reflected not just high averages but consistently high individual ratings across all satisfaction items. Tables 2 and 3 outline the standard scoring ranges for each metric. This combined quantitative-qualitative process ensured that the scoring system was both evidence-informed and grounded in practical experience and application, increasing its utility for program improvement and ease of internal replication.

Step 5: Implement framework and generate net effects diagrams

Workshop data was organized, separated by language, and analyzed using Microsoft Excel. Participant- and staff-reported data were compiled across all selected metrics and organized by RE-AIM subdomains (Tables 2 and 3). Raw metric values (e.g., percent change in knowledge, Likert-scale ratings) were converted

https://doi.org/10.1017/cts.2025.10190 Published online by Cambridge University Press

Table 2. (*Continued*)

RE-AIM Domain: Subdomain	Metric	Value (SD)	Standardized Scoring System	Standardized Metric Score	Subdomain Average Score
Effectiveness: Short-term goals	Participant change in knowledge	27.7% (0.55)	1 = <9.9% 2 = 10.0%-19.9% 3 = 20.0%-29.9% 4 = 30.0%-44.9% 5 = 45.0%+	3	3.5
	Participant perceived knowledge improvement	3.53 (0.65)	1 = 1.00-2.49 2 = 2.50-3.19 3 = 3.20-3.55 4 = 3.56-3.85 5 = 3.86+	3	
	Participant self-efficacy/ confidence levels	3.57 (0.65)	1 = 1.00-2.49 2 = 2.50-3.19 3 = 3.20-3.55 4 = 3.56-3.85 5 = 3.86+	4	
	Participant intention to apply workshop content in their own lives	3.62 (0.82)	1 = 1.00-2.49 2 = 2.50-3.19 3 = 3.20-3.55 4 = 3.56-3.85 5 = 3.86+	4	
Implementation: Evidence-based	Level of evidence-based content rating (from CE staff)	3.83 (0.45)	1 = 1.00-2.49 2 = 2.50-3.19 3 = 3.20-3.55 4 = 3.56-3.85 5 = 3.86+	4	4.0
Implementation: Responsiveness to community needs	Expert reflection/ CE staff perceptions - how did it go?	4.31 (0.67)	1 = 1.00-2.99 2 = 3.00-3.49 3 = 3.50-3.99 4 = 4.00-4.49 5 = 4.50+	4	3.5
	Level of cultural relevancy (from CE staff)	3.96 (0.85)	1 = 1.00-2.99 2 = 3.00-3.49 3 = 3.50-3.99 4 = 4.00-4.49 5 = 4.50+	3	
Effectiveness: Medium-term goals	Participant application of knowledge/skill	TBD	TBD	TBD	TBD
	Participant impact	TBD	TBD	TBD	

RE-AIM = Reach, Effectiveness, Adoption, Implementation, and Maintenance; SD = standard deviation; CE = Community Engagement; TBD = To Be Determined.

into standardized scores based on the validated scoring thresholds described in Step 4.

For subdomains composed of multiple metrics, we first calculated individual metric based standardized scores on a 1-to-5 scale. We then averaged them to create an overall subdomain score, allowing us to summarize performance across multiple indicators within each RE-AIM domain. This approach enabled consistent aggregation across both ELW and Spanish-language workshops (SLW) while maintaining sensitivity to nuances of each contributing metric.

All data processing was conducted using basic formulas and logical functions, which allowed for transparent calculations, ease of error checking, and replicability. We generated Net Effects Diagrams within Excel to visually compare the performance of ELW versus SLW across RE-AIM subdomains (Figure 2). By plotting subdomain scores for each group along the same axes, we could visualize how program groups differed from one another (e.g., higher "Reach: Attendance" but lower "Effectiveness: Short-term goals") and begin interpreting patterns to guide future programming.

https://doi.org/10.1017/cts.2025.10190 Published online by Cambridge University Press

Table 3. Operationalized outputs for Spanish-language workshops (January - September 2024)

RE-AIM Domain: Subdomain	Metric	Value (SD)	Standardized Scoring System	Standardized Metric Score	Subdomain Average Score
Reach: Frequency	Average monthly number of workshops offered	6.11 (2.42)	1 = 0.00-0.49 2 = 0.50-0.99 3 = 1.00-1.99 4 = 2.00-2.99 5 = 3.00+	5	5.0
Reach: Attendance	Perceived outreach effort by CE staff	4.53 (0.60)	1 = 1.00-2.99 2 = 3.00-3.49 3 = 3.50-3.99 4 = 4.00-4.49 5 = 4.50+	5	
	Average attendance per workshop	18.42 (9.70)	1 = 0.00-1.99 2 = 2.00-4.99 3 = 5.00-7.99 4 = 8.00-11.99 5 = 12.00+	5	5.0
	Perceived participant turnout by CE staff	4.44 (0.83)	1 = 1.00-2.59 2 = 2.60-3.19 3 = 3.20-3.79 4 = 3.80-4.39 5 = 4.40+	5	
	Characteristics of participants (actual in the context of recruited)	4.51 (0.74)	1 = 1.00-2.99 2 = 3.00-3.49 3 = 3.50-3.99 4 = 4.00-4.49 5 = 4.50+	5	
Effectiveness: Immediate reaction	Participant satisfaction	3.58 (0.64)	1 = 1.00-2.49 2 = 2.50-3.09 3 = 3.10-3.39 4 = 3.40-3.69 5 = 3.70+	4	4.0
	Perceived level of participant engagement by CE staff	4.11 (0.85)	1 = 1.00-2.59 2 = 2.60-3.19 3 = 3.20-3.79 4 = 3.80-4.39 5 = 4.40+	4	

(*Continued*)

Step 6: Interpret results and apply programmatic adjustments

We interpreted the Net Effects Diagram results to identify key patterns in program performance and used those insights to inform potential programmatic refinements. The Net Effects Diagrams that we generated provided a clear visual representation of differences in program implementation and outcomes between ELW and SLW, supporting both internal team reflection and broader stakeholder discussion.

Key Patterns and Insights:

- Reach: SLW had slightly higher subdomain scores for frequency of workshop delivery and participant attendance. This suggests greater consistency and participation compared to ELW.
- Effectiveness: SLW had higher scores in short-term effectiveness, including knowledge gain, perceived knowledge improvement, confidence in applying information, and intention to use the material, indicating a stronger immediate impact on participants in the SLW.
- Implementation: Scores for facilitator responsiveness to participant needs and adherence to evidence-based practices from SLW were higher, suggesting stronger alignment between delivery methods and participant expectations.

The combined results displayed in the Net Effects Diagram highlighted a pattern of stronger performance for SLW, particularly in the *Reach* and *Implementation* domains. These differences were discussed in depth during a CE team review session, where we used the visual outputs to identify areas for

https://doi.org/10.1017/cts.2025.10190 Published online by Cambridge University Press

Table 3. (*Continued*)

RE-AIM Domain: Subdomain	Metric	Value (SD)	Standardized Scoring System	Standardized Metric Score	Subdomain Average Score
Effectiveness: Short-term goals	Participant change in knowledge	32.6% (1.17)	1 = <9.9% 2 = 10.0%-19.9% 3 = 20.0%-29.9% 4 = 30.0%-44.9% 5 = 45.0%+	4	4.0
	Participant perceived knowledge improvement	3.59 (0.71)	1 = 1.00-2.49 2 = 2.50-3.19 3 = 3.20-3.55 4 = 3.56-3.85 5 = 3.86+	4	
	Participant self-efficacy/ confidence levels	3.58 (0.71)	1 = 1.00-2.49 2 = 2.50-3.19 3 = 3.20-3.55 4 = 3.56-3.85 5 = 3.86+	4	
	Participant intention to apply workshop content in their own lives	3.58 (0.73)	1 = 1.00-2.49 2 = 2.50-3.19 3 = 3.20-3.55 4 = 3.56-3.85 5 = 3.86+	4	
Implementation: Evidence-based	Level of evidence-based content rating (from CE staff)	4.12 (0.66)	1 = 1.00-2.49 2 = 2.50-3.19 3 = 3.20-3.55 4 = 3.56-3.85 5 = 3.86+	5	5.0
Implementation: Responsiveness to community needs	Expert reflection/staff perceptions - how did it go?	4.53 (0.71)	1 = 1.00-2.99 2 = 3.00-3.49 3 = 3.50-3.99 4 = 4.00-4.49 5 = 4.50+	4	4.0
	Level of cultural relevancy	4.56 (0.56)	1 = 1.00-2.99 2 = 3.00-3.49 3 = 3.50-3.99 4 = 4.00-4.49 5 = 4.50+	4	
Effectiveness: Medium-term goals	Participant application of knowledge/skill	TBD	TBD	TBD	TBD
	Participant impact	TBD	TBD	TBD	

RE-AIM = Reach, Effectiveness, Adoption, Implementation, and Maintenance; SD = standard deviation; CE = Community Engagement; TBD = To Be Determined.

adaptation in the implementation of ELW. Another adjustment identified was to better tailor the content for ELW, as we recognized the opportunity to further adapt materials to reflect the specific needs, experiences, and contexts of the participants. This might include revising examples, emphasizing relevant topics, and incorporating more community-specific language and examples.

Given that SLW were offered in more established community partner settings with long-standing relationships and ELW were part of a recent expansion into new communities, these findings aligned with team expectations. The use of the operationalized RE-AIM framework helped these nuances emerge and provided a clear path forward for continuous quality improvement and targeted program refinement. As these data-driven programmatic adjustments are implemented, the measurement process can be replicated, creating multiple Net Effects Diagram "snapshots" that may aid in understanding what drives changes in overall effectiveness over time.

https://doi.org/10.1017/cts.2025.10190 Published online by Cambridge University Press

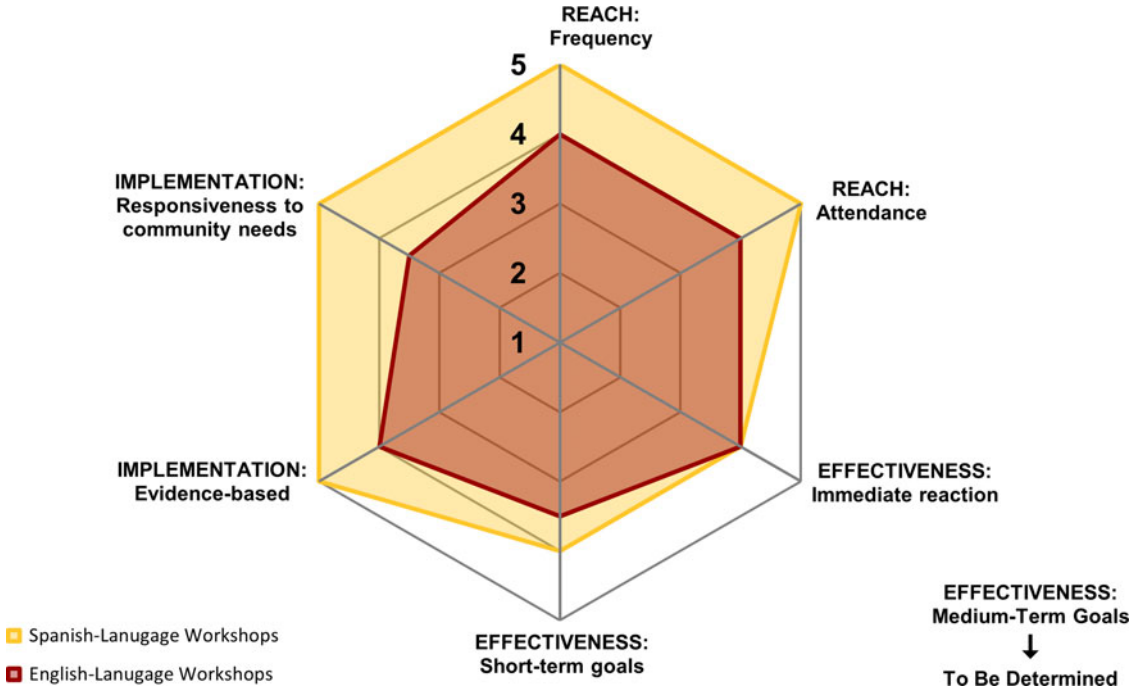


Figure 2. Net effects diagram comparing operationalized outputs for english-language workshops versus spanish-language workshops.

Step 7: Ongoing refinement of operationalized framework

Our team's next step will be to conduct a structured reflection process to assess the clarity, utility, and continued relevance of the current operationalized RE-AIM framework as applied to our health education workshops. This will include a focused review of how we defined each RE-AIM domain, the usefulness and interpretability of our selected subdomains, and the appropriateness of scoring thresholds used to standardize and compare outcomes. Our team will consider whether the current framework sufficiently captures what matters most to both community participants and partners. Key questions for discussion will include:

- Do the current domain and subdomain definitions remain aligned with how the workshops are evolving?
- Are any important aspects of workshop quality, accessibility, or impact missing from the current framework?
- Do the scoring thresholds accurately reflect meaningful differences in performance?

While the framework supported useful comparisons and decision-making, there may be areas where further refinement could improve clarity and alignment. For example, some subdomain metrics may require more precise definitions, and certain scoring thresholds may need to be recalibrated as more data become available. The future iterations of the framework may involve adjusting metric definitions or adding new subdomains based on emerging implementation priorities, revising score ranges to better reflect updated benchmarks or community expectations, and incorporating additional qualitative feedback loops from community participants and facilitators. Any changes made will include the rationale and implications for comparison over time, supporting transparency and continued usability. This refinement process will reflect our team's commitment to maintaining a responsive and evidence-informed evaluation approach that evolves alongside the program and the communities it serves.

Discussion

The paper presents and demonstrates a structured, replicable step-by-step evaluation approach grounded in the RE-AIM framework that supports rigorous, stakeholder-informed evaluation. Although originally designed for use in community-engaged research and implementation settings, it has the potential to be applied in other fields and settings. By integrating domain specific subdomains, scoring standardization, and visualization through Net Effects Diagrams, this approach allows for consistent assessment and actionable comparisons across diverse programs or delivery contexts. It also allows for data-driven assessments of what is working, where, and why, demystifying the challenge of answering, "how do we know?." The application of this method to our health education workshops demonstrated its feasibility and highlighted how the structured use of RE-AIM can reveal nuanced performance patterns and guide targeted program improvements and overall assessment of programming.

Structuring RE-AIM and creating comparable metrics

RE-AIM's broad domains can be difficult to apply in settings where data is: (1) sourced from multiple tools and levels; (2) varied in type; (3) short-term in nature; and/or (4) embedded in dynamic community contexts. Our approach addresses this by defining tailored subdomains, grounded in local implementation realities and refined through engagement with the team. This enables consistent monitoring while preserving the contextual nuances critical to CE work. In our workshops, this process helped translate abstract constructs like "Effectiveness" into specific measures, such as participant-reported confidence levels and increases in

https://doi.org/10.1017/cts.2025.10190 Published online by Cambridge University Press

knowledge. Our subdomains revealed key differences between ELW and SLW that would have been obscured by high-level metrics alone.

The development of standardized subdomain scores allowed for the comparison of various data types and from participants and facilitators across workshop topics, locations, and populations, an essential step for CE teams managing multiple programs. This scoring system offers a structured rubric system and a more meaningful understanding of the key differences between groups that accommodates program and implementation specific realities. For example, SLW consistently outperformed ELW in measures of "Reach: Attendance" and "Effectiveness: Short-term goals." Rather than interpreting this as a failure of ELW, our scoring system contextualized the results, prompting useful questions about outreach strategies and content relevance. For example, our team noted that many of the participants in SLW were reached through trusted community and faith-based partners. To improve *Reach* for ELW, the team recommended expanding recruitment efforts across our communities to include similar trusted venues. In this way, scoring supported both interpretation and action.

Visual tools for shared understanding and decision-making

The use of Net Effects Diagrams to visualize RE-AIM scores across subdomains further enabled accessible and data-informed decision-making. These tools helped our team quickly identify where programs were excelling, where adjustments or help may be needed to facilitate collaborative interpretation across roles and expertise levels. When we applied the diagram to compare groups, the visual differences helped generate immediate, understandable insights, such as the greater engagement observed in SLW and the need to revisit outreach strategies for ELW. This reinforces the value of clear, interpretable tools for facilitating feedback.

Limitations and considerations

Several limitations and considerations should be noted about this approach and opportunities for adoption. First, CE work often relies on self-reported outcomes that can be limited in capturing long-term behavior change and can introduce the potential for response bias. However, self-reported data is still an acceptable and commonly used data collection method to evaluate activities like health education workshops that have a focus on engagement, knowledge comprehension, and short-term outcomes. Future iterations of our CE evaluation will incorporate follow-up assessments to further understand medium-term outcomes.

Second, small sample sizes, particularly in subgroup comparisons, can limit generalizability. However, the approach's value lies in its ability to identify trends early and guide strategy adjustments, rather than to establish causal effects. Third, while the use of mean values as our measure of central tendency and the application of them for creating standardized scores allowed for more granular between-group comparison, this process can obscure variability within the data and may be sensitive to outliers. Future applications of this approach will consider alternative strategies such as median-based scoring, trimming outliers, or sensitivity analyses to improve robustness of comparisons. Others applying this approach should carefully review the distribution of their data and select central tendency measures that best fit their context and analytical objectives.

While this approach provides a structured process for identifying and defining domains, subdomains, and metrics, the development of rating scales and parameters remains context-specific and may vary across applications. While this flexibility allows for tailoring to program priorities and data availability, the approach is not intended to generate universally standardized measures and is best understood as a structured guide rather than a fixed standard. Future applications across multiple programs will refine common parameters to enhance comparability.

Importantly, this approach is not specific to workshops and can be applied to other CE activities where systematic monitoring and adaptation are needed (e.g., community-academic partnerships, resource navigation, or capacity building activities). Its blend of structure and flexibility makes it especially well-suited to the varied and evolving landscape of CE and other fields [17]. We also recognize that the reporting of this proof-of-concept represents its use for one type of activity administered by our team who also developed this operationalized approach. This approach has begun to be implemented elsewhere within our institute to help develop the formal and developmental evaluation of our institute's research trial recruitment dashboard. It has helped them, from a programmatic approach, to understand what metrics needed to be assessed to evaluate effectiveness over time.

Implications and future directions

This approach offers a guide for CE programs seeking to evaluate and improve their efforts systematically, rigorously, and comprehensively. Our application of Net Effects Diagrams shows how the operationalized RE-AIM framework can be practically implemented, interpreted, and adapted to inform meaningful improvements to CE program design and delivery. This ultimately supports more impactful and responsive engagement with the communities we serve.

Future directions include scaling this approach across our other CE activities, expanding outcome tracking to include longer-term measures, and integrating results into more dynamic database systems and interactive dashboards to enhance accessibility and responsiveness. As this approach expands beyond the proof-of-concept phase, we will continue to work collaboratively and incorporate more direct input from our stakeholders. To support evaluation capacity building across the CTSA network and beyond, this approach could be disseminated as a toolkit or adapted into training modules and interactive webinars. Such resources would enable broader uptake and customization of the method across settings while reinforcing shared learning and cross-site evaluation infrastructure.

Acknowledgements. The authors would like to acknowledge the contributions of the Community Health Workers/Promotoras de Salud on the Community Engagement team at SC CTSI, who delivered the workshops and provided vital feedback, the community members who participated in our workshops and completed surveys, and our community partners for their support and for providing their space and time for us to deliver our workshops.

The artificial intelligence (AI) tool ChatGPT (OpenAI) was used to provide editorial feedback in the editing and revision process of this manuscript. The authors used AI tools for tasks such as grammar correction and consistency. All content was reviewed, edited, and approved by the authors to ensure accuracy, originality, and alignment with the project objectives.

Author contributions. **Brian Do-Golden:** Conceptualization, Data curation, Methodology, Software, Visualization, Writing – original draft, Writing – review and editing; **Nicole Wolfe:** Conceptualization, Data curation, Project administration, Supervision, Writing – original draft, Writing – review and editing; **Nicole M. G. Maccalla:** Conceptualization, Methodology, Writing – review and editing; **James Settles:** Writing – original draft, Writing-review and

https://doi.org/10.1017/cts.2025.10190 Published online by Cambridge University Press

editing; **Michele D. Kipke:** Conceptualization, Supervision, Writing – review and editing.

Funding statement. This work was supported by grants UL1TR001855 from the National Center for Advancing Translational Science (NCATS) of the US National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Competing interests. The authors have no conflicts of interest to disclose.

References

1. **National Institutes of Health**. Clinical and translational science awards (CTSA) program, Published October 11, 2024. (https://ncats.nih.gov/research/research-activities/ctsa) Accessed January 22, 2025.
2. **National Institutes of Health**. PAR-24-272: clinical and translational science award (UM1 clinical trial optional). Published September 4, 2024. (https://grants.nih.gov/grants/guide/pa-files/par-24-272.html) Accessed January 22, 2025.
3. **Towfighi A, Orechwa AZ, Aragón TJ, et al.** Bridging the gap between research, policy, and practice: lessons learned from academic–public partnerships in the CTSA network. *J Clin Transl Sci.* 2020;**4**:201–208. doi: 10.1017/cts.2020.23.
4. **Concannon TW, Fuster M, Saunders T, et al.** A systematic review of stakeholder engagement in comparative effectiveness and patient-centered outcomes research. *J Gen Intern Med.* 2014;**29**:1692–1701. doi: 10.1007/s11606-014-2878-x.
5. **Eder MM, Evans E, Funes M, et al.** Defining and measuring community engagement and community-engaged research: Clinical and Translational Science institutional practices. *Prog Community Health Partnersh Res Educ Action*. 2018;**12**:145–156. doi: 10.1353/cpr.2018.0034.
6. **Eccles MP, Mittman BS.** Welcome to implementation science. *Implement Sci.* 2006;**1**:1. doi: 10.1186/1748-5908-1-1.
7. **Bauer MS, Damschroder L, Hagedorn H, Smith J, Kilbourne AM.** An introduction to implementation science for the non-specialist. *BMC Psychol.* 2015;**3**:32. doi: 10.1186/s40359-015-0089-9.
8. **Chambers DA, Glasgow RE, Stange KC.** The dynamic sustainability framework: addressing the paradox of sustainment amid ongoing change. *Implement Sci.* 2013;**8**:117. doi: 10.1186/1748-5908-8-117.
9. **Proctor E, Silmere H, Raghavan R, et al.** Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. *Adm Policy Ment Health*. 2009;**36**:65–76. doi: 10.1007/s10488-009-0237-7.
10. **Dolor RJ, Proctor E, Stevens KR, Boone LR, Meissner P, Baldwin LM.** Dissemination and implementation science activities across the Clinical Translational Science Award (CTSA) Consortium: report from a survey of CTSA leaders. *J Clin Transl Sci*. 2019;**4**:188–194. doi: 10.1017/cts.2019.422.
11. **Glasgow RE, Vogt TM, Boles SM.** Evaluating the public health impact of health promotion interventions: the RE-AIM framework. *Am J Public Health*. 1999;**89**:1322–1327. doi: 10.2105/ajph.89.9.1322.
12. **Glasgow RE, Battaglia C, McCreight M, Ayele RA, Rabin BA.** Making implementation science more rapid: use of the RE-AIM framework for mid-course adaptations across five health services research projects in the Veterans Health Administration. *Front Public Health*. 2020;**8**:194. doi: 10.3389/fpubh.2020.00194.
13. **Gunn CM, Sprague Martinez LS, Battaglia TA, et al.** Integrating community engagement with implementation science to advance the measurement of translational science. *J Clin Transl Sci.* 2022;**6**:e107. doi: 10.1017/cts.2022.433.
14. **Saary MJ.** Radar plots: a useful way for presenting multivariate health care data. *J Clin Epidemiol.* 2008;**61**:311–317. doi: 10.1016/j.jclinepi.2007.04.021.
15. **Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG.** Research electronic data capture (REDCap) – a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;**42**:377–381. doi: 10.1016/j.jbi.2008.08.010.
16. **Harris PA, Taylor R, Minor BL, et al.** REDCap consortium. The REDCap consortium: building an international community of software partners. *J Biomed Inform*. 2019;**95**:103208. doi: 10.1016/j.jbi.2019.103208.
17. **Luger TM, Hamilton AB, Ture G.** Measuring community-engaged research contexts, processes, and outcomes: a mapping review. *Milbank Q*. 2020;**98**:. 493–553. doi: 10.1111/1468-0009.12458.